

(19) 世界知的所有権機関  
国際事務局(43) 国際公開日  
2000年12月28日 (28.12.2000)

PCT

(10) 国際公開番号  
WO 00/79479 A1

(51) 国際特許分類<sup>7</sup>: G06N 3/00

(21) 国際出願番号: PCT/JP00/04050

(22) 国際出願日: 2000年6月21日 (21.06.2000)

(25) 国際出願の言語: 日本語

(26) 国際公開の言語: 日本語

(30) 優先権データ:  
特願平11/174426 1999年6月21日 (21.06.1999) JP

(71) 出願人(米国を除く全ての指定国について): 科学技術振興事業団 (JAPAN SCIENCE AND TECHNOLOGY CORPORATION) [JP/JP]; 〒332-0012 埼玉県川口市本町四丁目1番8号 Saitama (JP).

(72) 発明者; および

(75) 発明者/出願人(米国についてのみ): 北野宏明 (KITANO, Hiroaki) [JP/JP]; 〒350-0035 埼玉県川越市西小

仙波町2-18-3 Saitama (JP). 諸橋峰雄 (MOROHASHI, Mineo) [JP/JP]; 〒114-0024 東京都北区西ケ原1-27-3-1010 Tokyo (JP). 京田耕司 (KYODA, Koji) [JP/JP]; 〒227-0036 神奈川県横浜市青葉区奈良町2423-208 Kanagawa (JP). 濱橋秀互 (HAMAHASHI, Shugo) [JP/JP]; 〒211-0025 神奈川県川崎市中原区木月1566 佐藤ビル303号室 Kanagawa (JP).

(74) 代理人: 加古 進 (KAKO, Susumu); 〒170-0013 東京都豊島区東池袋一丁目20番2号 池袋ホワイトハウスビル Tokyo (JP).

(81) 指定国(国内): US.

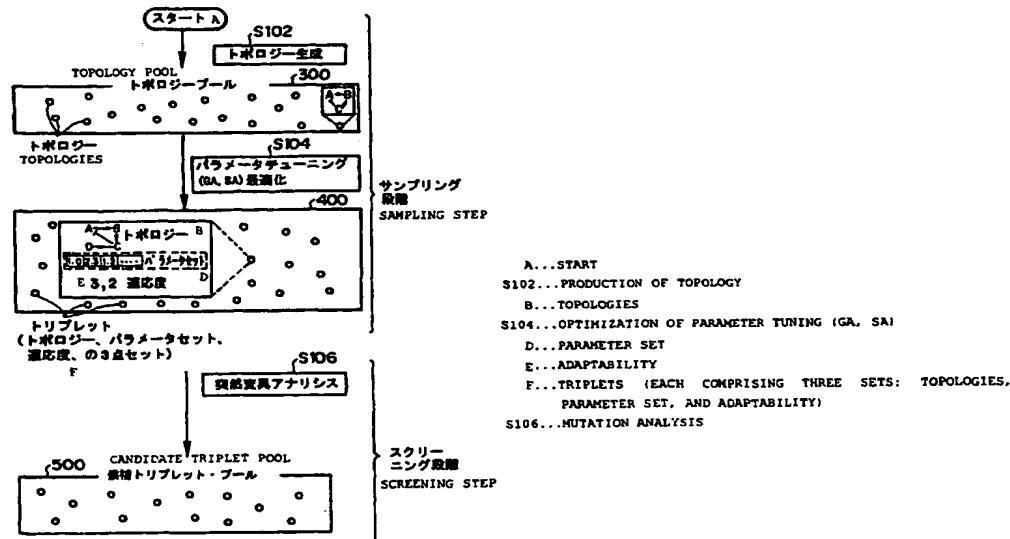
(84) 指定国(広域): ヨーロッパ特許 (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

添付公開書類:  
— 国際調査報告書

[統葉有]

(54) Title: METHOD AND DEVICE FOR NETWORK INFERENCE

(54) 発明の名称: ネットワーク推定方法および装置



WO 00/79479 A1

(57) Abstract: A network structure (topology) capable of reproducing a target profile is produced and a topology pool (300) is produced (S102). A parameter capable of best reproducing a target profile is determined for each produced network structure, and the errors are calculated (S104). Triplets being combinations of the topologies, the parameters, and errors are stored in a triplet pool (400). These operations are carried out at a sampling step. The triplets in the triplet pool (400) are screened to select triplets conforming to the object. For the screening, for example, mutation analysis is used. These operations are carried out at a screening step. (S106). Upon completion of the screening step, a candidate triplet pool (500) is produced.

[統葉有]



2文字コード及び他の略語については、定期発行される各PCTガゼットの巻頭に掲載されている「コードと略語のガイダンスノート」を参照。

---

(57) 要約:

ターゲット・プロファイルを再現する可能性のあるネットワーク構造（トポロジー）を生成して、トポロジー・プール300を作成する（S102）。生成したネットワーク構造のそれぞれに、ターゲット・プロファイルを最もよく再現することができるパラメータを決定する。また、その誤差も計算する。（S104）。トポロジーに対して、パラメータと誤差を組み合わせたトリプレットをトリプレット・プール400に格納する。ここまででは、サンプリング段階である。これから、目的に適合したトリプレットをトリプレット・プール400からスクリーニングする段階であり、このスクリーニングには、例えば、突然変異アナリシスを用いている（S106）。スクリーニング段階が終了すると、候補トリプレット・プール500ができる。

## 明細書

## ネットワーク推定方法および装置

## 技術分野

本発明は、互いに関連し、干渉し合う別々の要素のネットワークとしてモデル化できるシステムを、そのネットワークの挙動や状態を表す情報から逆に推定することに関する。特に、生物学的現象における遺伝子の発現データ、タンパク質の濃度データ等から、その現象を引き起こしている遺伝子の制御ネットワークや酵素やタンパク質の反応を示す代謝ネットワークの構造等を推定することに関する。

## 背景技術

最近の分子生物学の急激な進歩によって、主要な生物に関する多くの情報が蓄積されつつある。その典型例が、酵母、線虫、ショウジョウバエである。例えば、線虫 (*C. elegans*) では、その全細胞系譜の推定、神経系の結合の同定などがすでに行われており、昨年、DNAの全塩基配列の決定が発表された。酵母、大腸菌、マイコプラズマ等のモデル生物においても、DNAの全塩基配列がすでに決定されており、究極の目標であるヒトのDNAの全塩基配列も数年以内には終了すると思われる。しかしながら、配列が分かるということと、遺伝子やその機能、さらには遺伝子間の相互作用が分かるという間には、大きなギャップがある。

このような遺伝子の機能や遺伝子間の相互作用の決定は非常に困難な仕事である。現在、分子生物学の多くの研究は、注目する現象に関わる遺伝子およびその転写産物の特定と、それらのカスケードの決定に重点を置い

ている。このような遺伝子間の相互作用の理解を行う段階において、複雑な相互作用を人間の直感のみを用いて行うことは極めて難しい。これは、データが大量にあることやこれに関わる相互作用等の関係も色々と考えられるからである。我々は、この困難をコンピュータという計算能力をもった道具を導入することで乗り越えようとしている。

この考えを図式したのが第1図である。これは、コンピュータを利用する部分（バーチャル）と、実際に生物学的に実験する部分（リアル）との関係を示す図である。第1図において、まず、解明しようとする現象を特定すると、現在知られている生物学的知識（第1図 フェーズーI A）から、モデルをコンピュータ上に実装する（第1図 フェーズーI B）。また、すでに仮説が提示されている場合には、その仮説を実装する。実装されたモデルを使ってシミュレーションを行い（第1図 フェーズーI C）、観測されているデータとの整合性を検証する（第1図 フェーズーI D）。シミュレーションの結果と、観測データと一致しない場合は、2つの可能性が考えられる。まず、シミュレーションが正確でない場合である。これは、シミュレーションの精度をあげて、確実なものとすることで解決する。次に考えられるのは、モデルが不完全または間違っている場合である。この場合は知られている実験データに整合するモデルを構築する必要がある。

その後に、できるならば、このモデルを用いて、コンピュータ上で実験を行う（第1図 フェーズーII F）。フェーズIIの後、又はフェーズIIができないなら、直接、仮説的に決めたモデル等が正しいか実際に生物学的な実験を開始する（第1図フェーズーIIIおよびIV）。

このようにして、「生物のリバース・エンジニアリング」を行うことができる。

このようなコンピュータ上におけるシミュレーションにより、仮説的に

求めたモデルの絞り込みを必要としている分野は、上述の生物の分野に限らない。

### 発明の開示

本発明の目的は、コンピュータにより、仮説的に求めたモデルの絞り込みを効率良く行うことである。

例えば分子生物学においては、DNAチップやマイクロアレイ、さらには、PCRなどをを利用して得られた遺伝子産物やタンパク質の量の時間的变化に関するデータ、電気生理学的実験や電圧感受性色素などを用いた神経回路網の活動に関するデータ等から、その背後にある遺伝子やタンパク質、酵素の相互作用のネットワーク、および、神経細胞のネットワークで表されるモデルを、コンピュータを用いて推定することである。与えられたデータから考えられるモデルの総数は莫大な数である。従来は、これを人間の勘に頼ってあたりを付け、生物学的な実験により確かめていた。本発明は、与えられたデータから、類似したデータを生成する可能性のあるモデル（ネットワーク）を、コンピュータを用いて推定することにより、確からしいモデルの数をある程度まで絞り込もうとするものである。これは、上述の第1図に示した場合では、フェーズIに関するものである。

これらの相互作用のモデルの推定（ネットワークの推定）は、疾病の原因遺伝子の特定や、投薬の効果の推定等に必要である。また、未知の遺伝子や遺伝子産物の推定にも用いることができる。

上記目的を達成するために、本発明は、互いに関連している要素間の関係をネットワークとして記述でき、前記要素により生成された結果であるデータが与えられたときにおける、前記与えられたデータから、前記データを再現できる候補ネットワークを推定する方法であって、与えられた前記データを再現する可能性のあるネットワークの構造およびそれに対応す

るパラメータ・セットを作成し、複数の候補ネットワークを得るステップと、前記得られたデータとは異なる、ネットワークから生成できる他のデータを用いて、前記得られたネットワークから適切な候補ネットワークを絞り込むステップとを備えることを特徴とする。

これにより、互いに関連し、干渉し合う別々の要素のネットワークとしてモデル化できるシステムを、そのネットワークの挙動、状態、生成物等を表すデータから逆に推定することができる。

前記候補ネットワークを得るステップには、前記与えられたデータを再現可能なネットワーク構造が少ないとときは、その全ネットワーク構造を作成するステップを含むことができる。

前記候補ネットワークを得るステップは、前記与えられたデータを再現可能なネットワーク構造を選択するステップを含むこともできる。これは、パラメータを含めたネットワークが多数存在する場合に行われる。この選択するステップとして、例えば、ランダムにネットワーク構造を選択することにより行われる。

このような選択が行われた場合、前記与えられたデータとの誤差が少ないデータを再現するネットワーク構造を高頻度で作成するステップを含むこともできる。

これにより、与えられたデータを再現する可能性の高い候補ネットワークを選択する確率が高くなる。

また、前記与えられたデータとの誤差が少ないデータを再現するネットワーク同士の一部を組み替えて、ネットワーク構造を作成するステップを含むことができる。これは遺伝的アルゴリズム（GA）等を用いることにより実現することができる。

これらのネットワーク構造は、ネットワーク構造空間を広く粗く探索して得られている。

この広域探索を補うために、ネットワーク構造を得るために探索として、任意ネットワークの構造的近傍を探索して、ネットワーク構造を作成するステップを含むことができる。この局所探索には、例えば、シミュレーテッド・アニーリング（SA）を用いて行うことができる。

このようにネットワーク構造を広域探索および局所探索をハイブリッドで用いることにより、効率的に最適なネットワーク構造を選択することができる。

前記ネットワークを得るステップは、同一のネットワーク構造に対して、1個のパラメータ、若しくは、複数であって互いに異なるパラメータのセットを推定しており、これには、例えば、遺伝的アルゴリズム（GA）シミュレーテッド・アニーリング（SA）、及び／又は、山登り法等の最適化手法を用いて推定している。

これにより、与えられたデータを再現するための、ネットワーク構造に対する最適なパラメータを早く推定することができる。

このようにして得られた候補ネットワークを絞り込むために、与えられたデータとは別のデータを用いているが、このデータとして、ネットワークの一部が変異したネットワークが生成したデータを用いて行うことができる。

このようにして、最終的に、効率的に比較的少数の最適化された候補ネットワークを得ることができる。

上述のネットワーク推定方法を実行する装置および、上述のネットワーク推定方法をコンピュータ・システムに実現させることができるプログラムを格納した記録媒体も本発明である。

#### 図面の簡単な説明

第1図は、分子生物学におけるコンピュータ利用を説明する図である。

第2図は、遺伝子ネットワークの例を示す図であり、(a)は遺伝子ネットワークの表示例、(b)はネットワーク例を示す。第2図(b)で、A, B, C, Dは遺伝子による発現型である。

第3図は、発現プロファイルおよびそれに対応するネットワークを示す図である。(a)は野生型の発現プロファイル、(b)は3つのノードの例、(c)はB-発現プロファイル、(d)はC-発現プロファイルである。

第4図は、実施形態における処理例の概略フローチャートである。

第5図はネットワークのファンアウト制限を説明する図である。(a)はノード間の結合を示す。(b)はファンアウトの制約と確率との関係を示す。

第6図は、選択処理のフローチャートであり、第7図は、選択処理の手法を説明するための図である。第7図では、トポロジーはそれぞれ異なるようN個ランダムに選択され、1個のトポロジーについて、GA、SA等によるパラメータ・チューニングにより複数個のトリプレットを導出することを示す。パラメータ・セットは全て異なる値のものである(例えば、 $P_0^0 \neq P_0^1 \neq \dots \neq P_0^M$ )。

第8図は、SAによる再探索を説明する図である。図8ではトポロジー空間に、P個のトリプレットがあり、斜線部分がSAの対象となる部分を示している。

第9図は、候補の選別処理を説明する図であり、トポロジーがx個のノードで構成されている場合を示す。

第10図は、適応度の総合処理の例を説明する図であり、全ての適応度の和を各トポロジーの適応度とした例を示す。

第11図(a) (b)は、SAを行う場合の表記例および初期集団を示す図である。

第12図は、GAの適用例を示す図である。(a)はトポロジーのパラメ

ータを示し、結合荷重としきい値（斜線部分）とのリストで表されている。

（b）は第1世代の個体集団が突然変異により第2世代の個体集団を形成することを示す。

第13図は、具体的にGA処理を行っている例を示す図である。交叉（クロスオーバ）を（a）に、突然変異を（b）に示す。

第14図は、2次元GAを適用した例を示す図であり、2次元GAの交叉の例を示す。（b）は（a）の結合行列で示した例を図化したものである。

第15図は、SAによる近傍探索を説明するための例であり、（a）は遷移図、（b）は遷移確率の変化を示すグラフ、トポロジーを（c）に示す。

第16図は、突然変異による候補トリプレット選択の処理を示すフローチャートである。

第17図は、遺伝子・代謝ネットワークの例を示す図であり、（a）は遺伝子代謝ネットワークの記号例で、これを用いたネットワーク例が（b）である。

第18図は、ターゲット・プロファイルのデータ例を示す図であり、（a）が0次元空間データ、（b）が1次元空間データ、（c）が2次元空間データ、（d）が3次元空間データの例を示す。

#### 発明を実施するための最良の形態

本発明の実施形態を、図面を参照して詳細に説明する。

本発明は、例えば、特定の生物学的現象における遺伝子の発現データ、タンパク質の濃度データ等のプロファイルから、その現象を引き起こしている遺伝子の制御ネットワークや、酵素やタンパク質の反応を示す代謝ネットワークの構造と反応の強さや速度などを決定するいろいろな変数を推定する場合に適用することができる。この場合を例に以下説明する。

（遺伝子ネットワーク）

本発明の実施形態を、第2図に示すような遺伝子の相互関係を示しているネットワークを例に説明する。第2図(a)は、遺伝子をノードとして、その遺伝子が発現する物質等の発現に対して活性を示す関係と、抑制を示す関係を示している。それをネットワーク構造として図示したものが第2図(b)である。第2図(b)で、A, B, C, Dは遺伝子による発現型である。

さて、与えられたターゲット・プロファイル(発現プロファイル)に対する遺伝子のネットワークは、構成要素(DNA, RNA, タンパク質等)間の反応関係を示すネットワーク構造(トポロジー)、ネットワーク内部での反応や状態のモデルを記述する際に必要な変数の集合であるパラメータ・セット、および、ターゲット・プロファイルとの類似度を示す適応度の3つを単位とした、「トリプレット」で表すことができる。

まず、遺伝子のネットワーク構造の例について説明する。構成要素(ノード)間の反応規則は、簡単化すると、 $N \times N$ の結合重み行列W(パラメータ・セット)および結合行列C(ネットワーク構造)で表現される。結合重み行列Wの値wは、例えば、+1.0と-1.0の間の実数をとる。wの負の値は抑制を示し、正の値は活性を示す。結合行列Cでは、要素c<sub>ij</sub>の0と1の2値で、ノードpとノードqの間の結合関係を示す。0はノード間に結合関係がないことを意味し、1は結合関係があることを示す。また、Nノードを有する遺伝子ネットワーク構造のある時間tにおける発現状態は、N次元空間のベクトルx(t)で表現することができる。ベクトルxの各要素は、x<sub>i</sub>(t) (i = 1, ..., N)で表される。そして、s<sub>i</sub>(t)を、要素i (i = 1, ..., N)の発現型の濃度であり、h<sub>i</sub>を活性要素iの活性レベルを決定するしきい値であるとすると、各要素は、例えば、次のように定義することができる。この関数は、様々に定義することが可能である。

$$x_i(t+1) = F \left( \sum_{j=1}^N c_{ji} w_{ji} s_i(t) - h_i \right)$$

ここで  $F(u) = \begin{cases} 0 & (u < 0) \\ u & (0 \leq u < 1) \\ 1 & (u \geq 1) \end{cases}$

#### (実施形態の処理の概略)

このような遺伝子ネットワークの場合は、複数の発現物質の濃度の時間的变化で示される発現プロファイルが与えられて、この発現プロファイルを発現する遺伝子ネットワークの候補を探し出すことが本発明の目的である。例えば、第3図(a)に示すような発現プロファイルを与えて、第3図(b)に示すような遺伝子ネットワークの候補を探し出すのである。なお、この第3図(a)に示した発現プロファイルは、野生型の遺伝子構造に対する発現プロファイルである。最初に用いる発現プロファイルは、これに示されるような野生型であることが望ましい。

ネットワークの結合(リンク)部分の数字は、上述における結合重み行列Wの値wの絶対値を示している。なお、結合重み行列Wの値wの符号は、結合の活性か抑制かにより正負が決められる。また、遺伝子Bがない突然変異(ミュータント)のプロファイル、遺伝子Cがない突然変異(ミュータント)のプロファイルがそれぞれ第3図(c)、第3図(d)に示されている。このプロファイルの用いられ方については後で説明する。

本実施形態は、第4図の概略フローチャートに示すようなプロセスにより行われる。第4図において、まず、対象である発現プロファイルを実現することができるネットワークの構造を想定し、そのネットワーク構造(トポロジー)を生成して、トポロジー・プール300を作成する(S102)。

想定されるネットワーク構造の数が少ない場合は、想定される全てのネットワーク構造を生成して、トポロジー・プール300に格納することができるが、想定されるネットワーク構造の数が多い場合は、トポロジー・プール300に格納するトポロジーに対して何らかの制限を設ける必要がある。

なお、必ずしも、トポロジー・プール300にトポロジーを生成して格納する必要はない。想定されるネットワーク構造により構成されるトポロジー空間を定義することができれば、この定義されたトポロジー空間から任意のトポロジーを生成するアルゴリズムを用いることにより、トポロジー・プール300からトポロジーを取り出すのと同様のことができる。

生成したトポロジーのそれぞれに、発現プロファイルからのパラメータを適応度を利用して適合させる(S104)。パラメータは、上述の結合重み行列Wの値としきい値である。また、適応度は、このパラメータの確からしさ(どれだけ合致しているか)である。この適応度は、評価関数を用いて計算されるが、例えば、この評価関数として最小自乗誤差などがある。

ネットワーク構造(トポロジー)に対して、パラメータと適応度を組み合わせられたものは、前述したように、トリプレット(トポロジー、パラメータ、適応度の3点セット)という。処理結果のトリプレットをトリプレット・プール400に格納する。トポロジーを選択してパラメータを適合させ、適応度を算出するときに、GA(遺伝的アルゴリズム)やSA(シミュレーテッド・アニーリング:焼きなまし法)等の手法を用いて、与えられている発現プロファイルに対して最適化する。

ここまででは、処理対象である標本(トリプレット)を抽出する段階である。これから、目的に適合した標本(トリプレット)をトリプレット・プール400から絞り込む段階となる。この処理には、例えば、突然変異ア

ナリシスを用いている (S106)。この処理により、標本 (トリプレット) が絞り込まれて、目的とする候補トリプレットの集合 (候補トリプレット・プール 500) ができる。

(ネットワーク抽出過程)

ネットワーク構造の生成 (S102) の処理は、例えば、最大でいくつのノードを含むネットワークを想定するか定める。そのネットワークの構造を、予備知識無しから生成するか、部分的にその構造が分かっているネットワークを基に生成する。生成したネットワークをトポロジー・プール 300 に格納する。この場合に用いられるネットワークの分かっている構造には、例えば、ファンアウトの制限がある。これを説明するのが第5図である。ファンアウトとは、ノードから他のノードへの働きかけ (活性、抑制) の数である。第5図 (a) は、A から B～E へ、4 個のファンアウトがあることを示している。このファンアウトの数には制約があると考えられ、この制約は、第5図 (b) に示す例のように、確率で示されている。第5図 (b) では、ファンアウトの数の確率は、2 付近で極大となり、5 以上のファンアウトの数の確率は無視できる程度であることを示している。なお、この制約も様々に定義することが可能である。

また、すでに同定されている結合や未結合が分かっている場合は、その結合以外の部分に対してネットワーク構造を発生させる。ネットワーク構造を発生させるときに、ある部分の結合構造を仮定して、発生させることも可能である。このような色々な構造に対する制約的な規則を適用して、トポロジーを作成することもできる。

トポロジー・プール 300 に格納されているネットワーク構造 (トポロジー) の全てにパラメータや適応度を適合して、トリプレット・プール 400 に格納するか、抽出して格納するかは、生成され得るネットワークの数による。生成され得るネットワークの数が少ないと、すなわち、ネット

トワークのノードの数が少ない場合は、あり得る全てのネットワークを網羅的に処理することが可能であるので、生成されたトポロジーに全てパラメータおよび適応度を適合させてトリプレット・プール400に格納する。しかし、生成されるネットワークの数が膨大で、全数を検討することができない場合は、トポロジーを選択してパラメータや適応度を割り当て、トリプレット・プール400に格納する。

(トポロジーが多数存在する場合)

さて、トポロジーの数が膨大で、トポロジーの選択が必要な場合における処理を、第6図のフローチャートと第7図および第8図を用いて、もっと詳しく説明する。

第7図および第6図のフローチャートを参照する。まず、最初にいくつのトポロジーを選択するか定める。この場合は、N個のトポロジーをランダムに選択するとする。そして、第4図のトポロジー・プール300（又は、トポロジー空間）から、ランダムに1つトポロジー310を取り出す(S202)。これに対して、与えられた発現プロファイルを実現できる発現パラメータを、GA（遺伝的アルゴリズム）／SA（シミュレーテッド・アニーリング）などを用いて適合させる(S204)。このとき、各発現パラメータによる発現プロファイルとの適応度の高いものから、異なるパラメータ・セットをM個(412, 414, 416, 418)とり、トリプレット・プール400に格納する。これをN個のトポロジーについて繰り返す。繰り返しにおいて、ランダムにトポロジーを選択するとき(S202)、前に選択したトポロジーとは異なる構造のものを選択する。繰り返しが終了した段階(S206でYES)では、N×Mのトリプレットがトリプレット・プール400に格納されている。

なお、上述では、ランダムにトポロジーを選択しているが、他の確率的サンプリングを用いてもよい。また、異なるパラメータ・セットをM個と

るとき、パラメータ・セット同士の類似度を計算して、その類似度が極めて近いとき（あるしきい値より小さいとき）は、同じとして選択しないようにもよい。これは、同じような解の近傍を複数選択しても意味が無いことが多いので、代表値のみを選択していることを意味している。

次に、トリプレット・プール400に格納されているN個のトリプレットのトポロジー組みかえをしたN個のネットワーク構造（トポロジー）をGA/SA等により得る（S208）。これらのN+1～2Nの得られた各トポロジーに対して、上述と同様に、パラメータ・セットをGA/SAにより適合させ、発現プロファイルとの適応度の高いものをM個残して、トリプレット・プール400に格納する（S210）。これにより、2N×M個のトリプレットがトリプレット・プール400に格納されることになる。この2N×M個のトリプレットから、適応度が高いものから所定のしきい値以下の値を有するP個を選択してトリプレット・プール400に残す（S212）。これで、トリプレット・プール400に格納されているトリプレットはP個となった。ここまで処理は、トリプレットに対する広域探索である。

#### （局所探索）

トリプレット・プール400に格納されているP個のトリプレットの近傍を、例えばSA（シミュレーテッド・アニーリング）により探索して、より適応度の高いトリプレットを探す（S214）。これを図示したのが、第8図である。第8図において、P個の各トリプレットにおけるトポロジー空間のSAの対象となる近傍に対して、探索を行っていることを示している。これにより、より適応度の高いトリプレットを得ることができる。より適応度が高いトリプレットが見つかった場合は、それもトリプレット・プール400に格納する（S216）。これにより、P個以上のトリプレットがトリプレット・プール400に格納されることになる。

なお、このとき、より適応度が高いトリプレットが、あるトリプレットの近傍に発見された場合、そのトリプレットに代えて、より適応度が高いトリプレットをトリプレット・プール400に格納することもできる。この場合、トリプレット・プール400に格納されているトリプレットの個数は、P個のままである。

ここで、SAを用いて、近傍の探索を行ったが、他の手法、例えば山登り法（可能性のある探索点のなかで最も有望な点を選んで探索を進めていく方法）等により行うことが可能である。

このようにして、トリプレット・プール400に格納されているトリプレットの適応度を高めているので、サンプリング段階を終了した時点で、トリプレット・プール400に格納されているトリプレット中に、目的とするトリプレットが入っていることが、高い確率であり得る。しかし、この段階のトリプレットの数は、まだ、これに対して生物学的な実験を行うのに、十分には小さくない。

#### （絞り込み過程）

次に、トリプレット・プール400に格納されているトリップレットから、より少数の候補トリプレットを選別する処理の段階である。この処理を、突然変異分析を用いて行っている第9図および第10図を例として説明する。

第9図において、トリプレット・プール400から各トリプレットに対して、ミュータント（突然変異）・トリプレットを作成する（S402）。この場合のミュータント・トリプレットは、ある遺伝子を除去（ノックアウト）して、その遺伝子からの結合を全て取り去ることで作成している。このとき、トリプレットにおけるその他の結合や、パラメータは変化させない。これを遺伝子（ネットワーク構造のノード）に対して行う。したがって、トリプレットのトポロジーにおけるノードのうち、x個の遺伝子の

## 1.5

それぞれのノードをノックアウトすると、ミュータント・トリプレットをそれぞれ格納したミュータント・プール452～456はx個できる。そして、与えられたそれぞれのターゲット・データである突然変異プロファイルとの間の適応度を各ミュータント・プールごとに評価する (S404)。そして、各ミュータント・プールごとに評価した適応度を統合して、最終的に、ある値以上の適応度を有する候補群を選出して、候補トリプレット・プール500に格納する (S406)。この場合、適応度により、得られた候補トリプレットをソートしておいて、その適応度を参照して、高い順に生物学的な実験を行ってもよい。

ここで用いる突然変異としては、上述のノックアウトの他に、例えば、ヘテロザイゴート (Heterozygote)、過剰発現 (over expression)、温度感受性突然変異 (temperature sensitive mutation) 等があり、いずれも使用可能である。例えば、ヘテロザイゴートは、相同染色体の特定の遺伝子座においての一方の対立遺伝子をつぶしてしまうことにより、全体としてその遺伝子の発現量が例えば半分などに減少してしまう突然変異である。また、過剰発現は、ある遺伝子について過剰に発現してしまう突然変異である。温度感受性突然変異は、温度感受性のある物質を用いて、計測中のある時点で遺伝子の発現を止めてしまう、又は、過剰に発現させる突然変異である。

#### (適応度の統合の計算例)

この適応度の統合の計算例について、第10図を用いて説明する。第10図では、全ての突然変異ごと（突然変異ではない野生型も含む）の各トポロジーごとの適応度を合計したものを、そのトポロジーの適応度として取り扱う例を示している。この計算例では単純な合計をしているが、必要に応じて、突然変異ごとに重み付けて加重加算等を行ってもよい。

なお、突然変異の発現プロファイルは、対象の全ての突然変異遺伝子ネ

ツトワークに対して得られているとは限らない。この場合は、得られている突然変異の発現プロファイルのみを用いて、適応度の算出および統合を行う。このようにして、最終的に、候補トリプレットを得ることができる。

(実施例)

具体的な遺伝子のトリプレットを例に、上述の処理を詳しく説明する。ここでは、第3図(a)に示したような時系列データである発現プロファイルを実現する遺伝子のネットワークを例とする。

まず、トポロジーの結合関係を示す結合行列をもとに染色体にコーディングを行う。ここでは、第3図(a)に示した発現物質が3つであるので、ノードが3つあるネットワークのトポロジーとしては、例えば、以下のようないくつかの結合行列Cで表される。これは、第3図(b)に示されているものである。

$$C = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & 0 & 0 \end{pmatrix}$$

このトポロジーの結合は4カ所あり、各々に対する重み付けをする必要があるので、4つの遺伝子型 (genotype) を持つ染色体を生成する。さらに各ノードのしきい値も最適化する必要があるので、結局、全体として4 (結合数) + 3 (ノードのしきい値数) で7の遺伝子型を持つ染色体を生成することになる。この場合の染色体の荷重およびしきい値の取りうる範囲 (実数値) は、第11図(a)のようになる。このような第3図(a)に示す発現プロファイルを実現可能な複数のトポロジーを発生して、トポロジー・プール300 (第4図) に格納する。

この発生した各トポロジーに対して、GAを用いてパラメータ (荷重お

よびしきい値) の最適化を行う。まず、トポロジーに対して決められた範囲のパラメータをN組ランダムに発生させ、染色体の初期集団とする。第11図(b)に、上述の結合行列で定められたトポロジーに対する初期集団を示す。このように、トポロジーに対して、リストの形で染色体が定まる。この初期集団に対してGAを適用することにより、最適化を行うのである。この最適化の指標として、各パラメータの適応度を用いる。

初期集団の適応度は、染色体のパラメータをもとに、各個体で発現プロファイルを生成し、ターゲット・プロファイルとの間で、例えば最小自乗誤差TSS (Total Sum Square error) の処理を行うことにより求める。この最小自乗誤差TSSは、以下の計算を行う。

第3図(a)に示すように、ターゲット・プロファイルが濃度の時系列データとして与えられている。このため、各発現物質A, B, Cそれぞれのある時間tにおける濃度は $L_{TA}(t)$ ,  $L_{TB}(t)$ ,  $L_{TC}(t)$ と表され、生成した染色体の各パラメータをもとに、上述の式(1)により計算される各発現物質の濃度は、 $L_{EA}(t)$ ,  $L_{EB}(t)$ ,  $L_{EC}(t)$ と表すことができる。これらの間の最小自乗誤差TSSは次のようにあらわすことができる。

$$TSS = \sum_{X}^{A,B,C} \sum_{t=0}^T (L_{TX}(t) - L_{EX}(t))^2$$

このようにして求めた最小自乗誤差TSSを各発生したネットワーク(染色体)の適応度とする。

さて、このように求められる適応度を利用することにより、GAを用いてトポロジー毎にパラメータ(結合荷重としきい値)の最適化を行うことができる。この処理について、第12図、第13図を用いて、GAのクロ

スオーバと突然変異を用いて最適化を行っている例で説明する。

第12図(a)で示されているように、トポロジーのパラメータは、結合荷重としきい値のリストで表されている。まず、親となる個体対の決定を、例えばトーナメント戦略を用いて行う。これは、トーナメント・サイズの数の個体(例えば2)をランダムに選択する。この選択した個体の中から一番適応度が高い個体を一方の親とする。これをもう一度行うことにより、親となる個体対が決定される。この個体対に対して、第12図(b)に示すように、交叉(crossover)を行って子孫のパラメータを作成し、さらに、突然変異を適用して、生成したパラメータで適応度を計算する。このような処理を、例えば、定められた数の子孫が作成されるまで行うと、第2世代の個体集団が得られる。

交叉および突然変異の処理を、第13図を用いて具体的に説明する。第13図(a)において、リストで表された親1および親2のパラメータの矢印で示された部分610を入れ替えて、子1および子2が作成される。これが交叉の処理である。第13図(b)は、子1の丸を付けられたパラメータの1つ(結合荷重0.4)620が、別の値(結合荷重0.6)630に変化したことが示されている。この変化が突然変異で、例えば、固定された確率でランダムにおきるように処理される。

この交叉と突然変異の処理を、例えば、決められた世代数分繰り返すことで、あるトポロジーに対するパラメータを最適化することができる。なお、全世代で、エリート保存戦略を適用してもよい。エリート保存戦略は、集団中で最も適応度の高い個体を、交叉や突然変異により変化させず、そのまま次世代に残す手法である。

以上の処理をすべてのトポロジーに対して終了したときに、その中で、適応度(この例の場合、最小自乗誤差)以下の適応度を持つトポロジーのサブグループD<sup>+</sup>を決定する。このサブグループD<sup>+</sup>は、次の式により定義

される。

$$D^+ = \{D^+ \in D \mid TSS(D^+) \leq T^+\}$$

抽出されたトポロジー・サブグループのトリプレット（トポロジー、パラメータ、適応度）を、トリプレット・プール400へ格納する。なお、この実施例ではトポロジー・プール300に格納されているすべてのトポロジーに対してGAにより最適なパラメータを作成しているが、トポロジー・プール300に格納されたトポロジーが莫大な数である場合には、定められた数のトポロジーをランダムに選択して、これに対して最適なパラメータをGAにより生成してもよい。ここまでが、トリプレットの広域探索である。

このトリプレット・プールに格納されているトリプレットに対して、さらに別のGA手法を適用して最適化したものを追加する。ここで適応するGAは2次元GAである。この2次元GAについて第14図を用いて説明する。第14図(a)は、結合行列で表した例で、それを図式化したものを第14図(b)に示す。第14図(a)に示すように、親1および親2の結合行列の十字形で示される部分（即ちBを中心とする結合関係：第14図(b)参照）を交叉させて、子1および子2を生成する。このように、第13図(a)で説明したリスト（パラメータ）上の交叉とは異なり、ネットワークの物理的な構造自体に着目して交叉を行っている。このような処理は、2次元構造の最適化に有用である。2次元GAを適用して得られたトポロジーに対してパラメータを、GAを前述と同様に用いることで最適化を行って求める。これを例えれば、決められた世代数実行する。そして、上述と同じように、適応度  $T^+$  以下であるトリプレットのみを、さらにトリプレット・プール400に格納する。この追加して行うGAでもエリー

ト保存戦略を適用することができる。

トリプレット・プール400に格納されているトリプレットを対象としてSA（シミュレーテッド・アニーリング）を実行して、各トポロジーの近傍トポロジー空間を探索し、得られたトポロジーに対して、前と同様にGAにより最適化されたパラメータを求める。この近傍の探索を行うのは、GAが局所的な探索能力に劣っているためである。

ここで用いているSAの基本アルゴリズムは、疑似コードを用いると、以下のようになる。これは、各トポロジーに対して適用される部分である。

```
main() { initialize()
    while(停止条件不適合) {
        while(平衡条件未到達) {
            次の状態生成()
            if (ACCEPT(E, E', T))
                更新()
        }
        T= REDUCE(T)
    }
}
```

上述のACCEPT(E, E', T)中のEは、SAではエネルギーと呼ばれる状態を規定する変数で、この場合は、各ネットワークの適応度を使用している。E'は、「次の状態生成()」処理で生成されたトリプレットの適応度である。また、Tは、SAのアルゴリズムでは温度パラメータで、この場合は、この処理のみで用いられる状態を規定する変数である。変数Tは、処理が始まるときに、初期値 $T_{init}$ に設定される。また、停止条件や平

衡条件は、例えば、処理回数を用いることができる。

「ACCEPT(E, E', T)」, 「REDUCE(T)」は、次のように定義される。

$$ACCEPT(E, E', T) = \begin{cases} 1 & E' \leq E \\ e^{-(E-E')/T} & \text{その他} \end{cases}$$

$$REDUCE(T) = \gamma T \quad (0 < \gamma \leq 1)$$

上述の疑似コードで表される処理を説明する。処理対象のトポロジーが選択されると、変数TとEとが初期化される。そして、回数で規定されるループの内側の繰り返し処理で、まず、「次の状態発生()」という処理を行う。この「次の状態発生()」は、以下のような処理を行う。

処理対象のトポロジーを基に、結合関係を変化させたトポロジーを発生させる。この変化は、第15図(a)に示した遷移図に基づき、処理対象のトポロジーを、確率Pにより変化させる。この遷移確率は、上述の変数Tを用いて、次のように定義される。

$$P = \begin{cases} T/T_{init} & T/T_{init} \leq \text{しきい値} \\ \text{しきい値} & \text{その他} \end{cases}$$

ここで、Tは上述の変数、しきい値Thresholdは、ここでは例えば、0.5に設定されている。しきい値を導入した理由は、遷移確率をあまり高くすると、対象としたトポロジーに対して多数の変化したトポロジーを探索することになり、近傍探索とはならないからである。Tは、外側のループの「REDUCE(T)」で変化するので、内側のループ内では遷移確率は変化しない。外側のループのステップにより、遷移確率がどのように変化するのかを示したのが、第15図(b)である。この場合、γは例えば、0.8

である。トポロジーの変化を第15図(c)に示したトポロジーを例に説明する。これは、第3図(b)のトポロジーと同じものである。なお、遷移確率を0.5とした場合で説明する。

第15図(c)において、AからAの結合は、活性化で結合しているので、第15図(a)の状態遷移図では、活性606の状態にある。この状態からは、未結合604への状態遷移のみが可能である。この状態遷移が起こる確率は0.5である。次に、AからBへの結合を見るところも活性として結合しているので、同様に、0.5の確率で、未結合の状態に変化する。BからBへの結合は、未結合なので、抑制と活性とに0.25づつ(合計で0.5)の確率で遷移する。以下同様に、すべての結合関係について、第15図(a)に示した状態遷移図と、第15図(b)の確率に基づいて、変化させる。

このようにして、対象のトポロジーを構成する各結合関係を上述で説明した確率で、各遷移図に従って変化させた結果得られたトポロジーに対して、次に、前述のGAを用いて、最適なパラメータを決定し、そのパラメータに対する適応度を計算する。これで新しいトリプレットが作成される。これで「次の状態発生()」は完了し、次の処理である「ACCEPT(E, E', T)」を実行する。

この「ACCEPT(E, E', T)」では、処理対象であるトポロジーを有するトリプレットの適応度(E)と、発生したトポロジーの適応度(E')を比較し、発生させたトポロジーの適応度が低い(より適合している)と、1(真)を返す。また、発生した適応度が高くても(適合していなくても)、ある確率で1(真)を返している。この場合、Tが高い場合(あまり処理回数が多くないとき)に1(真)を返す確率を多くしている。「ACCEPT(E, E', T)」は1(真)を返すと、「更新()」処理が行われて、「次の状態発生()」で発生したトリプレットに更新する。

このようにして、内側のループを平衡状態が得られるまで（この場合はある回数の処理が行われるまで）続ける。所定回数の処理が行われると、外側の処理（「REDUCE(T)」）が行われて、新しいTが定められ、また、このTを用いて、内側の処理が行われる。外側の処理も、この場合、例えばある所定回数行われる。

そして、最終的にたどり着いたトリプレットの適応度がT<sup>+</sup>であると、トリプレット・プールに入れる。

上述のようにこの段階終了時では、トリプレット・プール400には、各トポロジーの近傍を探索して、より良い適応度のトリプレットと置き換えられたものが格納されている。このトリプレットから、最終的な候補トリプレットを選択する処理を以下で説明する。これは、種々の突然変異の発現プロファイルを用いて、これに対する適応度を計算することにより行う。この処理を、第16図のフローチャートを用いて説明する

これには、まず、第3図(c)、第3図(d)に示されているような、突然変異によるプロファイルを用意する。第3図(c)、第3図(d)は、遺伝子Bがない突然変異(ミュータント)のプロファイル、遺伝子Cがない突然変異(ミュータント)のプロファイルである。そして、それぞれに対応したミュータント・プールを用意し、上述のトリプレット・プール内の全トリプレットを各ミュータント・プールに格納する(S802)。

各ミュータント・プールに格納されたトリプレットに対して、トリプレットのトポロジーをその突然変異のトポロジーとし、パラメータはそのまま用いて、対応する突然変異のプロファイルに対する適応度を計算する(S804)。

各ミュータント・プールにおけるトリプレットのうち、計算した適応度が各突然変異ごとの定めた所定のしきい値以下であるトリプレットを選択する(S806)。選択した各ミュータント・プール内のトリプレットで、

全部のミュータント・プールに共通するトリプレットであり、かつ、所定のしきい値以下の適応度を有するものに対応するトリプレットを、候補トリプレット・プールへ格納する (S808)。

これで、適切な候補トリプレットを確定することができる。

#### (他の適用)

上述の説明では、遺伝子ネットワークの例で説明したが、例えば遺伝子・代謝回路ネットワークにも適用することが可能である。第17図は、遺伝子・代謝回路のネットワークについて示している。第17図 (a) に示すように、遺伝子と、代謝における酵素、タンパク質をノードとして、活性、抑制、媒介等の関係をリンクで表す。そして、これを用いたネットワーク例を第17図 (b) に示す。この第17図 (b) に示したようなネットワーク構造に対して、上述で説明した処理を適用する。

上述においては、対象ネットワークとして、遺伝子ネットワーク、代謝回路ネットワークを示したが、そのほかにも神経回路ネットワーク等に対しても適用することができる。この場合のデータ (ターゲット・プロファイル) としては、ニューロン活動電位等を用いる必要がある。

このように、本発明の手法は色々なものに対して用いることができる。適用できるターゲット・プロファイルの色々な例を第18図に示す。第18図 (a) は、ターゲット・プロファイルが0次元空間データである場合を示している。このデータの例としては、時間や周波数に対して変化するようなデータである。例えば、濃度、活性度、電位等がある。第18図 (b) は、1次元空間データを示している。データは、空間として  $x$  に、および、時間又は空間に依存する量である。この量としては、濃度、活性度、電位等がある。第18図 (c) および (d) として示したターゲット・プロファイルは、2次元空間データおよび3次元空間データである。これは、2次元 ( $x, y$ ) および3次元 ( $x, y, z$ ) の空間、および、

時間又は周波数に依存する量のデータである。この量としては、濃度、活性度、電位等がある。第12図に示すように、色々なデータ形式で、ターゲット・プロファイルを与えることができる。

このように、ターゲット・プロファイルが存在し、そのターゲット・プロファイルを生じるネットワーク構成を推定するような場合に、本発明は有効である。

本発明は、スタンド・アローンのコンピュータ・システムばかりではなく、複数のコンピュータから構成される例えばクライアント・サーバ・システム等に適用してもよい。

本発明に関するプログラムを格納した記憶媒体から、プログラムをシステムで読み出して実行することにより、本発明の構成を実現することができる。この記録媒体には、DVD、MD、MO、フロッピー・ディスク、CD-ROM、磁気テープ、ROMカセット等がある。

#### 産業上の利用可能性

上記の説明のように、本発明は、データとしてターゲット・プロファイルが存在し、そのターゲット・プロファイルを生じるネットワーク構成を推定するような場合に、極めて有効である。

特に、遺伝子ネットワークや代謝ネットワークのように、分子生物学におけるモデルの推定に対して有効に適用することができる。

## 請求の範囲

1. 互いに関連している要素間の関係をネットワークとして記述でき、前記要素により生成されたデータが与えられたときにおける、前記与えられたデータから、前記データを再現できる候補ネットワークを推定する方法であって、

与えられた前記データを再現する可能性のあるネットワークの構造およびそれに対応するパラメータ・セットを作成して、複数の候補ネットワークを得るステップと、

前記得られたデータとは異なる、ネットワークから生成できる他のデータを用いて、前記得られたネットワークから適切な候補ネットワークを絞り込むステップと

を備えることを特徴とするネットワーク推定方法。

2. 請求の範囲第1項に記載のネットワーク推定方法において、前記候補ネットワークを得るステップは、前記与えられたデータを再現可能な全てのネットワーク構造を作成するステップを含むことを特徴とするネットワーク推定方法。

3. 請求の範囲第1項または第2項に記載のネットワーク推定方法において、前記候補ネットワークを得るステップは、前記与えられたデータを再現可能なネットワーク構造を選択するステップを含むことを特徴とするネットワーク推定方法。

4. 請求の範囲第3項に記載のネットワーク推定方法において、前記候補ネットワークを得るステップは、前記与えられたデータとの誤差が少ないデータを再現するネットワーク構造を高頻度で作成するステップを含むことを特徴とするネットワーク推定方法。

5. 請求の範囲第3項または第4項に記載のネットワーク推定方法におい

て、前記候補ネットワークを得るステップは、前記与えられたデータとの誤差が少ないデータを再現するネットワーク同士の一部を組み替えて、ネットワーク構造を作成するステップを含むことを特徴とするネットワーク推定方法。

6. 請求の範囲第3項～第5項のいずれかに記載のネットワーク推定方法において、前記候補ネットワークを得るステップは、任意ネットワークの構造的近傍を探索して、ネットワーク構造を作成するステップを含むことを特徴とするネットワーク推定方法。

7. 請求の範囲第1項～第6項のいずれかに記載のネットワーク推定方法において、前記候補ネットワークを得るステップは、同一のネットワーク構造に対して、1個のパラメータ、又は、複数であって互いに異なるパラメータのセットを推定することを特徴とするネットワーク推定方法。

8. 請求の範囲第7項に記載のネットワーク推定方法において、前記候補ネットワークを得るステップは、遺伝的アルゴリズム(GA)、シミュレーテッド・アニーリング(SA)、及び／又は、山登り法等の最適化手法を用いてパラメータを推定するステップを含むことを特徴とするネットワーク推定方法。

9. 請求の範囲第1項～第8項のいずれかに記載のネットワーク推定方法において、前記他のデータは、ネットワークの一部が変異したネットワークが生成したデータであることを特徴とするネットワーク推定方法。

10. 互いに関連している要素間の関係をネットワークとして記述でき、前記要素により生成されたデータが与えられたときにおける、前記与えられたデータから、前記データを再現できる候補ネットワークを推定するネットワーク推定装置であって、

ネットワーク構造と、それに対応するパラメータ・セットとで構成されるネットワークを記憶する第1の記憶手段と、

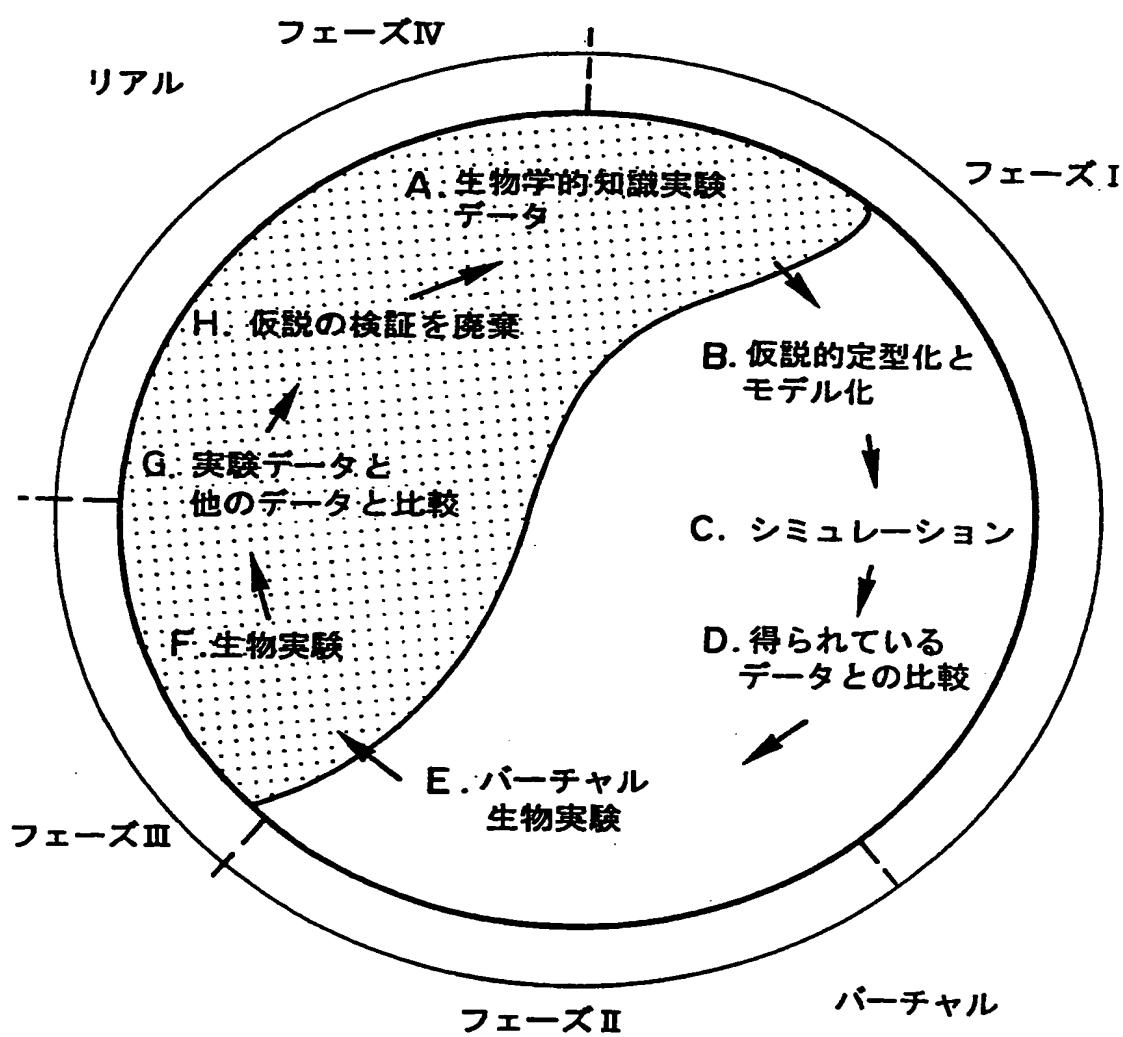
最終的な候補としてのネットワーク構造と、それに対応するパラメータ・セットとを格納する第2の記憶手段と、

与えられた前記データを再現する可能性のあるネットワークの構造およびそれに対応するパラメータ・セットを作成し、複数の候補ネットワークを得て、前記第1の記憶手段に格納する手段と、

前記得られたデータとは異なる、ネットワークから生成できる他のデータを用いて、前記第1の記憶手段に格納されているネットワークから適切な候補ネットワークを絞り込み、前記第2の記憶手段に格納する手段とを備えることを特徴とするネットワーク推定装置。

11. 請求の範囲第1項～第9項のいずれかに記載のネットワーク推定方法をコンピュータ・システムに実現させることができるプログラムを格納した記録媒体。

## 第 1 図

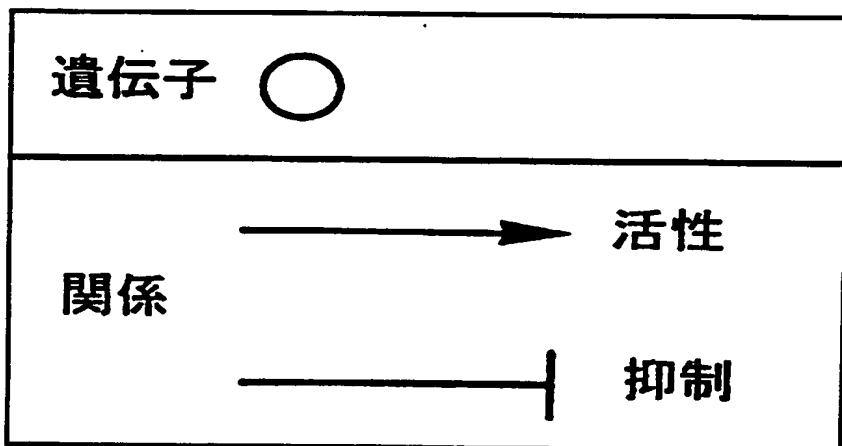


**THIS PAGE BLANK (USPTO)**

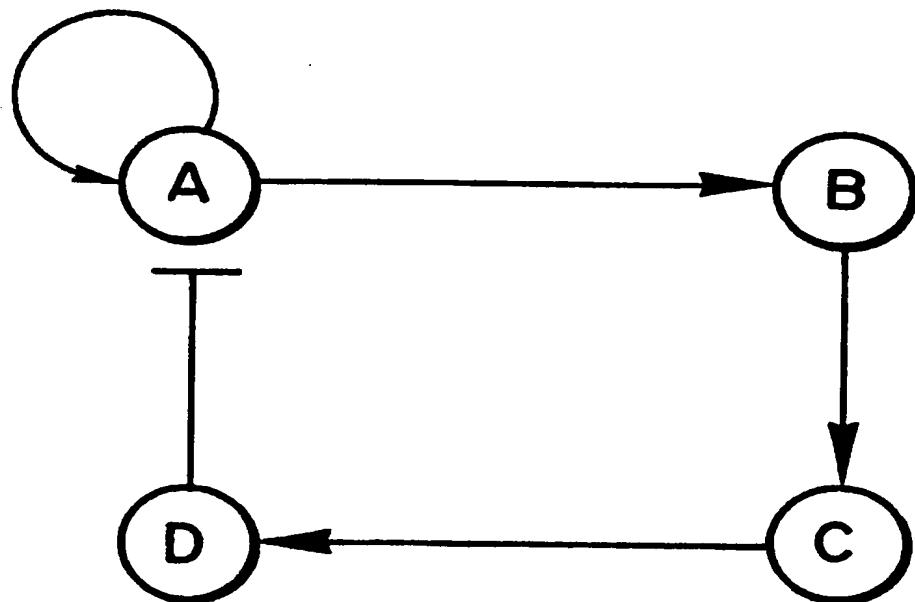
2 / 18

## 第 2 図

(a)

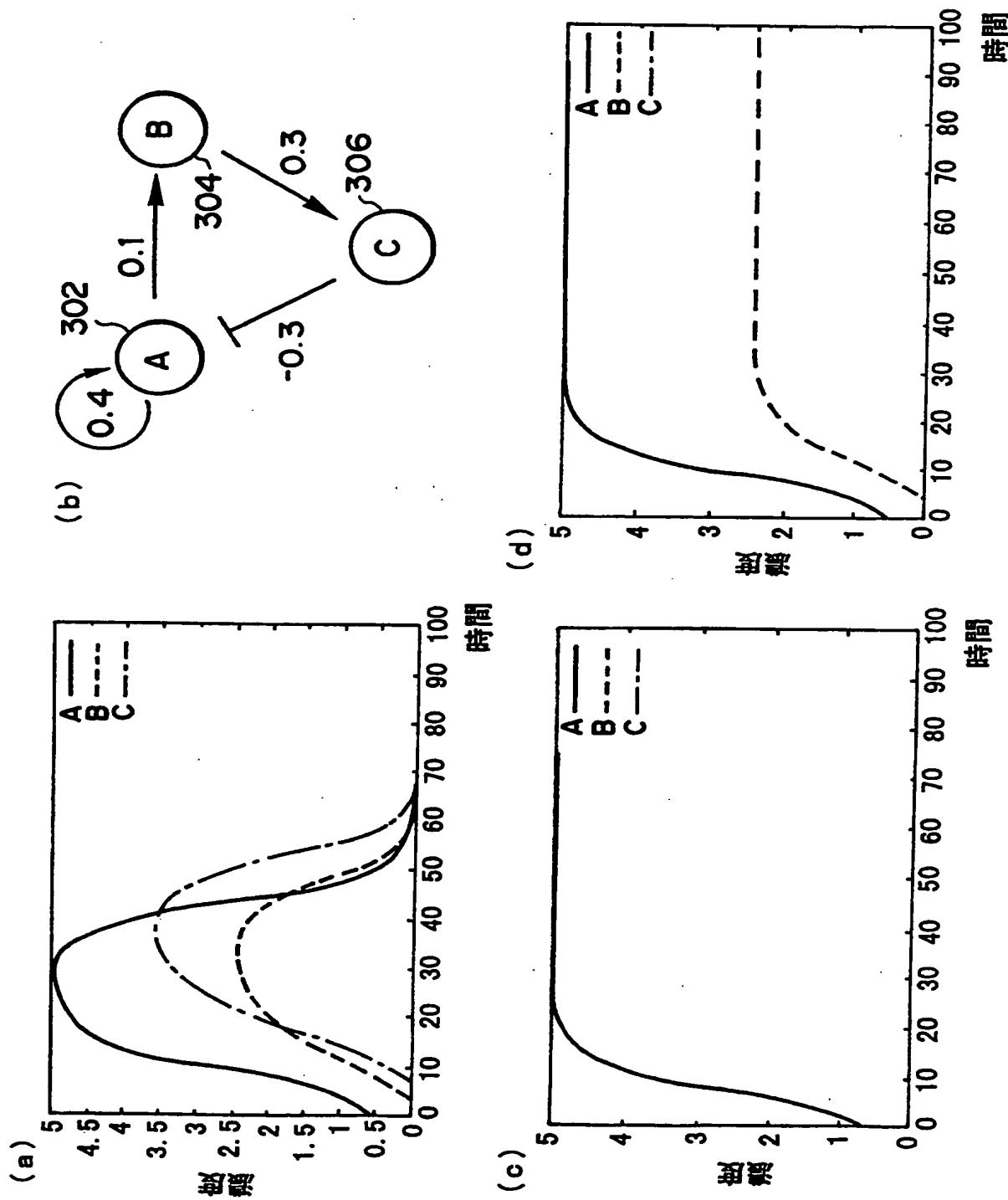


(b)



**THIS PAGE BLANK (USPTO)**

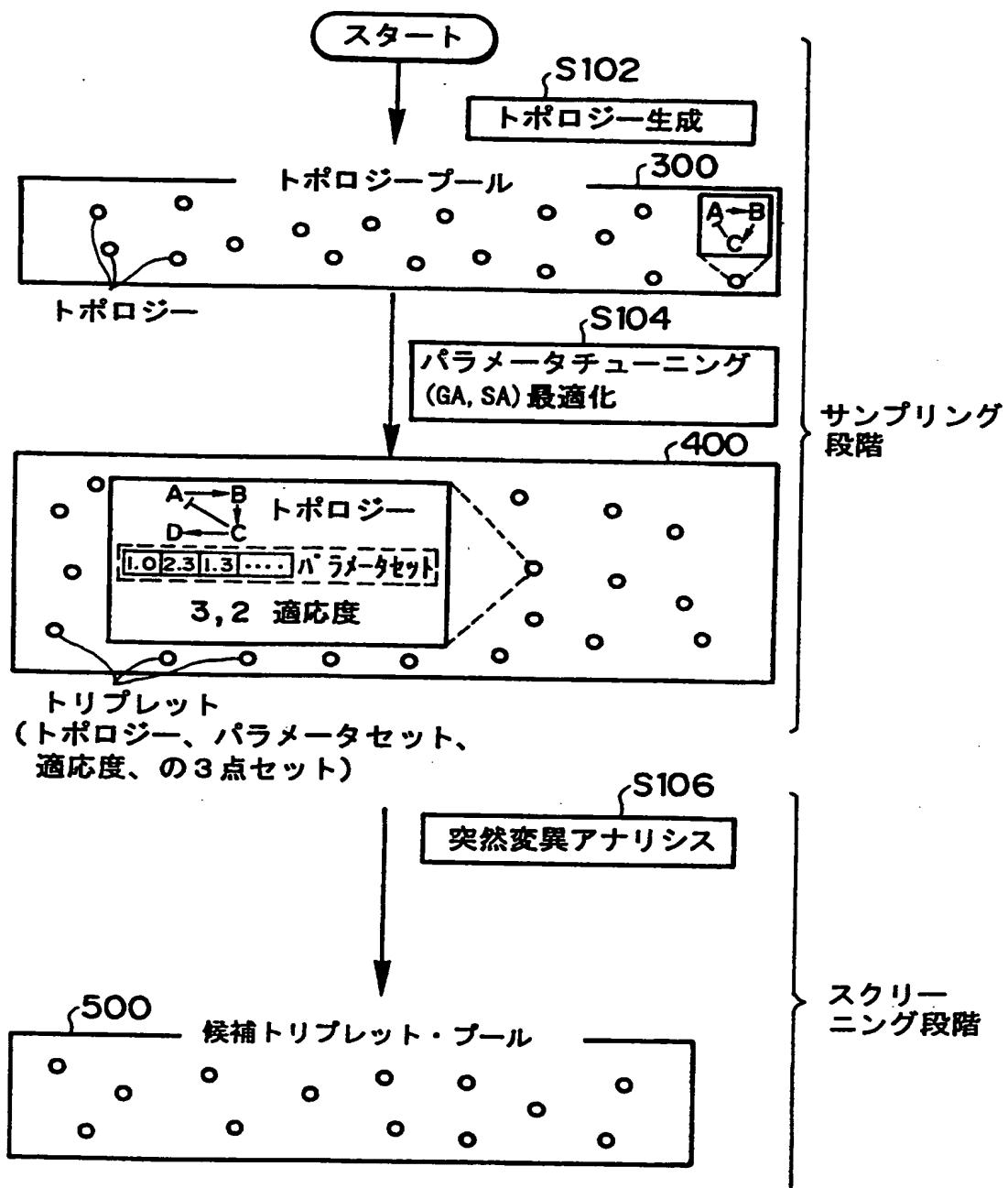
## 第 3 図



**THIS PAGE BLANK (USPTO)**

4 / 18

第 4 図

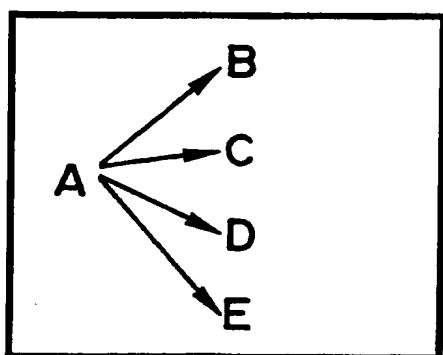


**THIS PAGE BLANK (USPTO)**

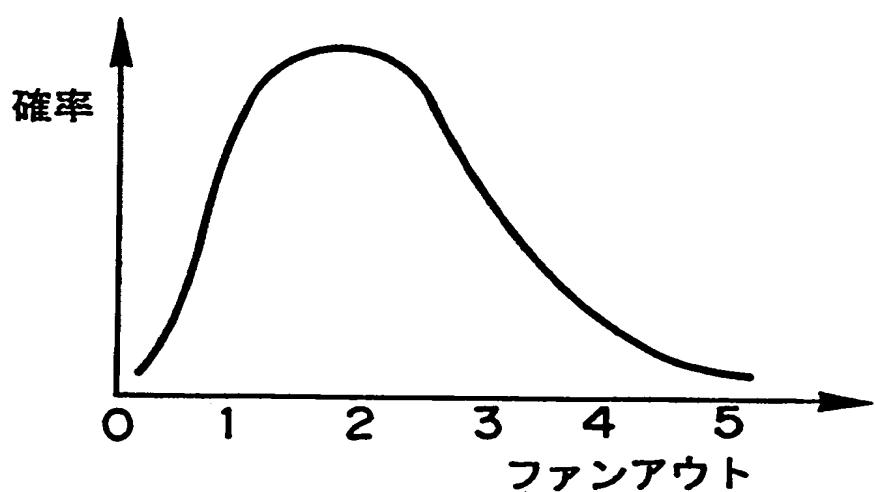
5 / 18

## 第 5 図

(a)



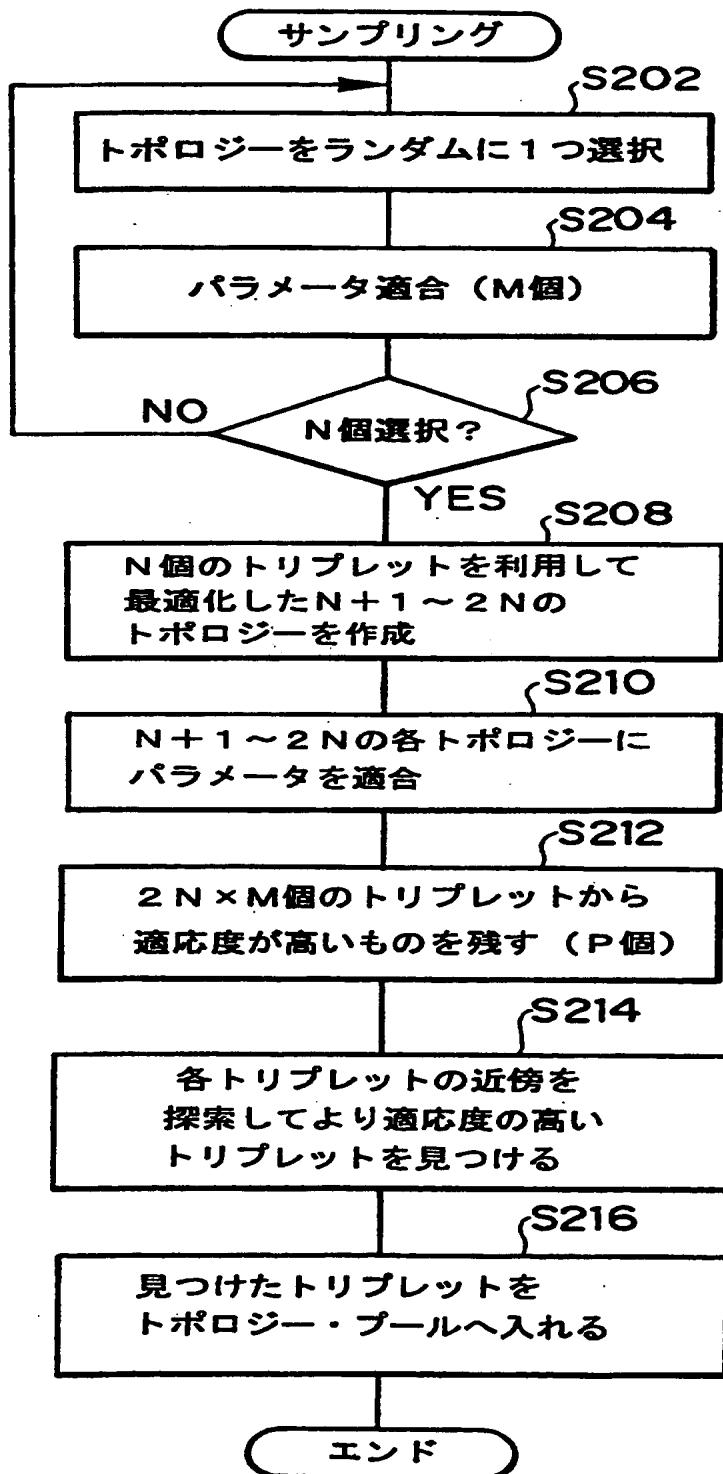
(b)



**THIS PAGE BLANK (USPTO)**

6 / 18

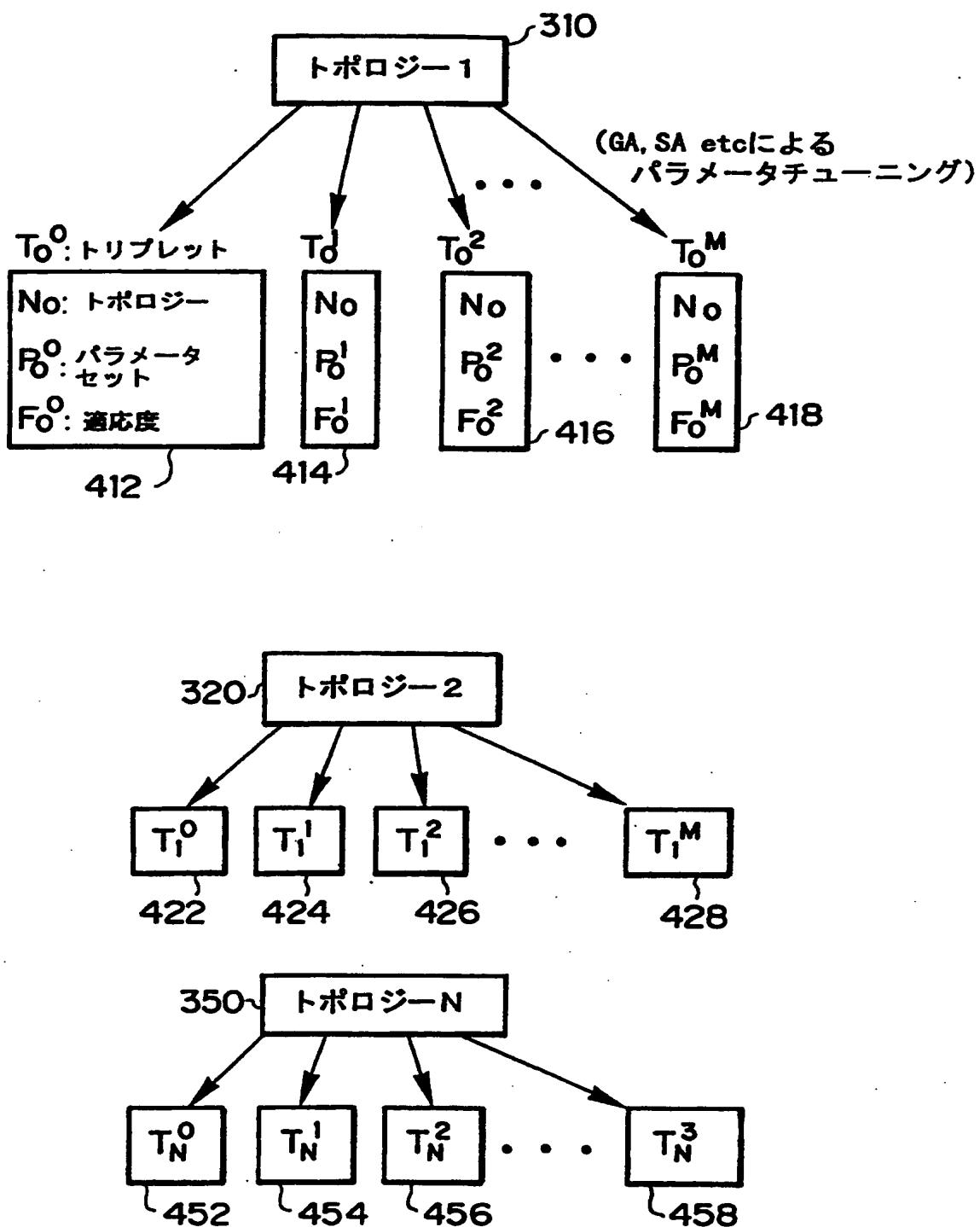
## 第 6 図



**THIS PAGE BLANK (USPTO)**

7 / 18

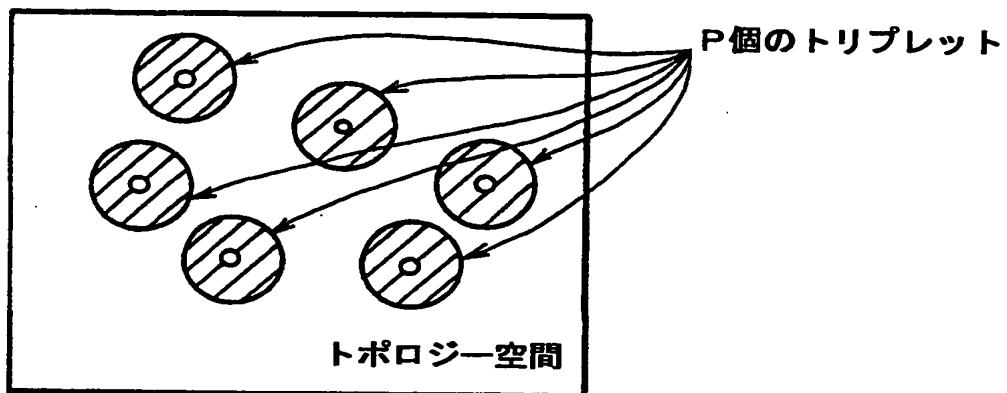
第 7 図



**THIS PAGE BLANK (USPTO)**

8 / 18

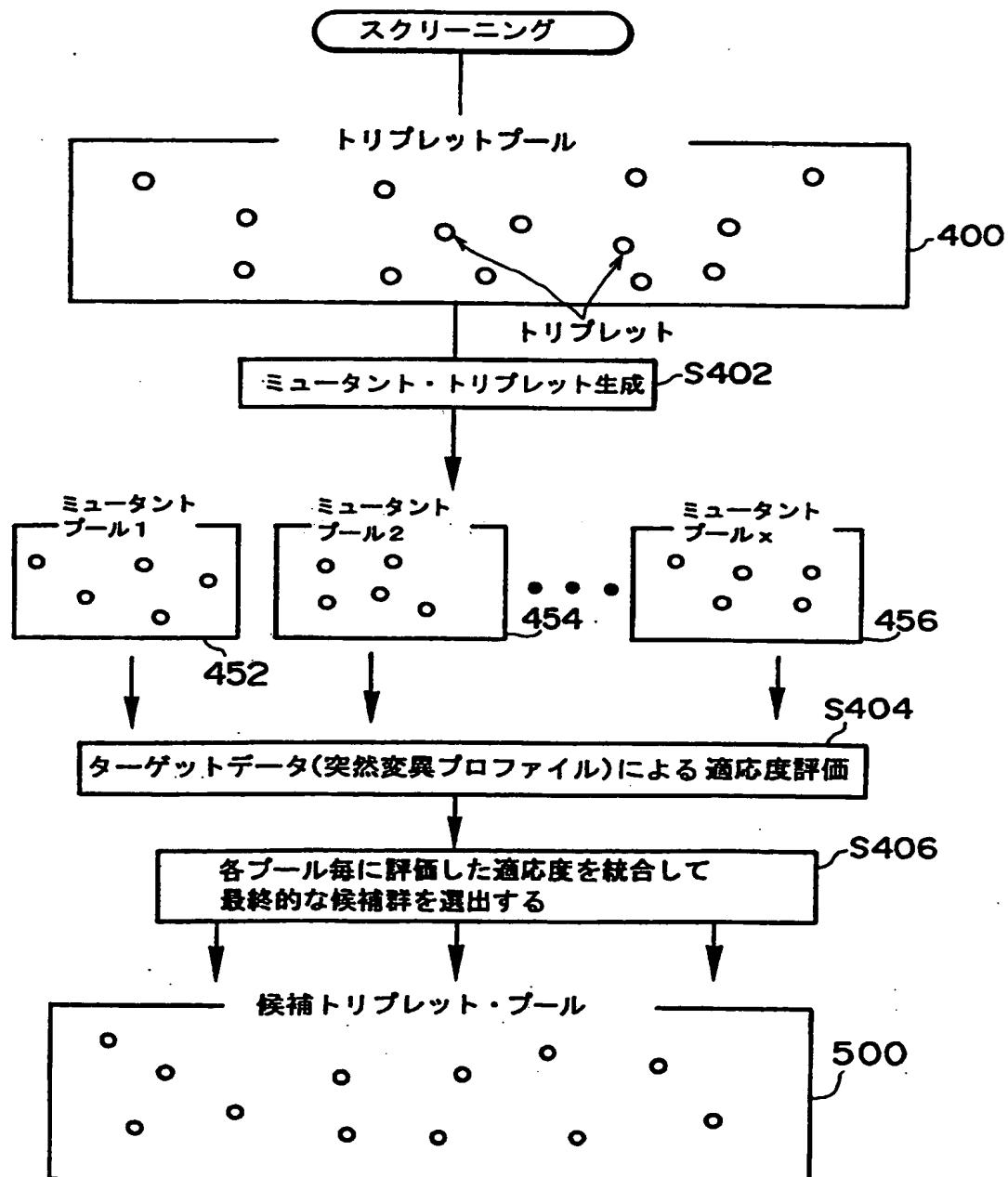
## 第 8 図



**THIS PAGE BLANK (USPTO)**

9 / 18

## 第 9 図



**THIS PAGE BLANK (USPTO)**

10/18

## 第 10 図

	トポロジー 1	トポロジー 2	...	トポロジー x
野生	12	15	...	32
突然変異 1	8	6	...	23
突然変異 2	16	10	...	12
突然変異 x	13	5	...	22
+				
	382	293		512

**THIS PAGE BLANK (USPTO)**

11/18

## 第 11 図

染色体

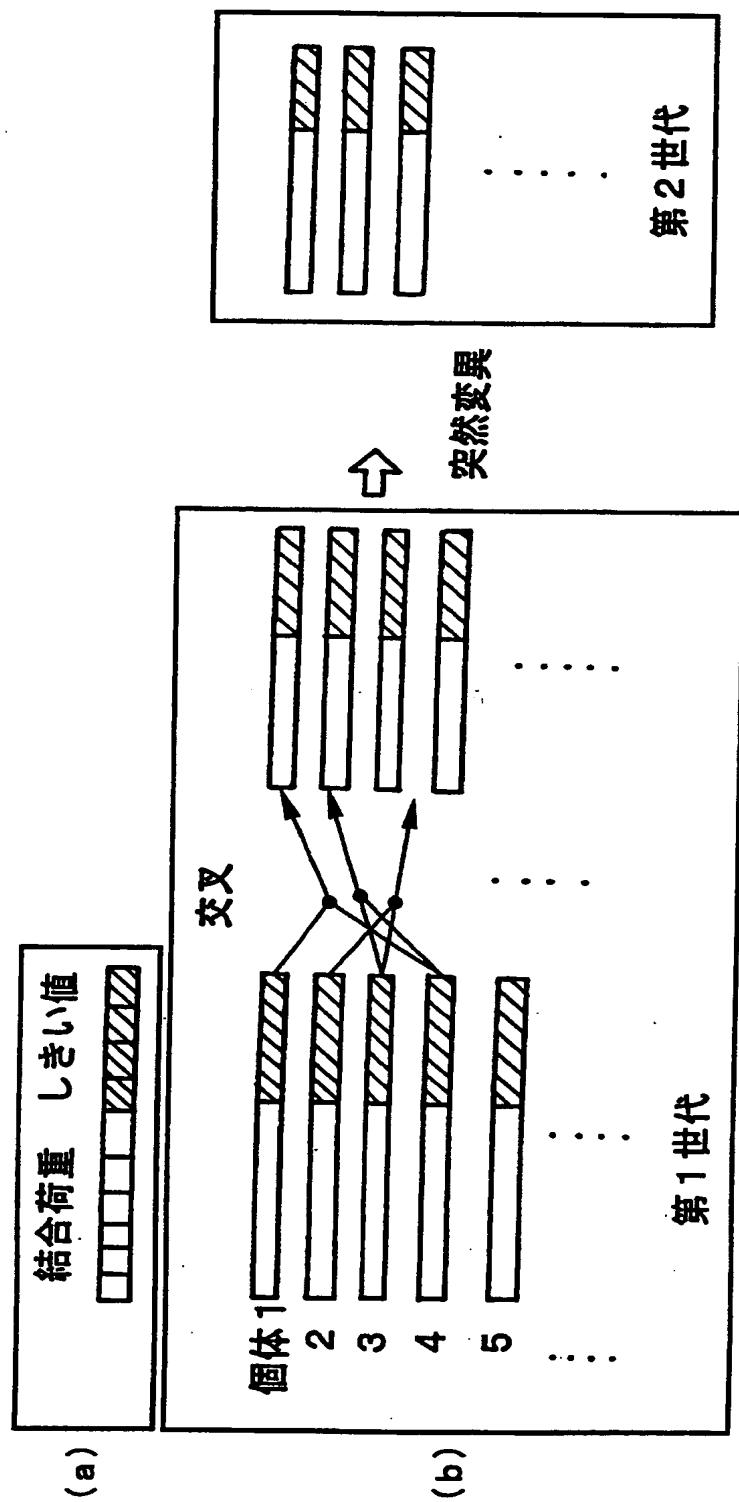
(a)	結合荷重				各ノードのしきい値		
	A→A	A→B	B→C	C→A	A	B	C
	[0~1]	[0~1]	[0~1]	[-1~0]	[0~1]	[0~1]	[0~1]

(b)

初期集団	0.2	0.4	0.2	-0.1	0.6	0.2	0.3	個体 1
	0.3	0.7	0.2	-0.3	0.6	0.2	0.4	個体 2
	0	0.2	0.5	-0.6	0.3	0.3	0.3	個体 3
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	0.3	0.2	0.6	-0.5	0.2	0.4	0.6	個体 N

**THIS PAGE BLANK (USPTO)**

## 第 12 図



**THIS PAGE BLANK (USPTO)**

13/18

## 第 13 図

(a)

					610
親 1	0.2 0.4	0.3 -0.2 0.3			0.2 0.5
親 2	0.1 0.2	0.2 -0.6 0.3			0.5 0.5
↓					
子 1	0.2 0.4	0.2 -0.6 0.6			0.2 0.5
子 2	0.1 0.2	0.3 -0.2 0.3			0.5 0.5

(b)

子 1	0.2	0.4	0.2 -0.6	0.6	0.2	0.5
		620				
		↓				
子 1'	0.2	0.6	0.2 -0.6	0.6	0.2	0.5
		630				
		↓				

**THIS PAGE BLANK (USPTO)**

## 第 14 図

(a)

親 1 の結合行列

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & -1 \\ 1 & 1 & 0 \end{pmatrix}$$

親 2 の結合行列

$$\begin{pmatrix} 0 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix}$$

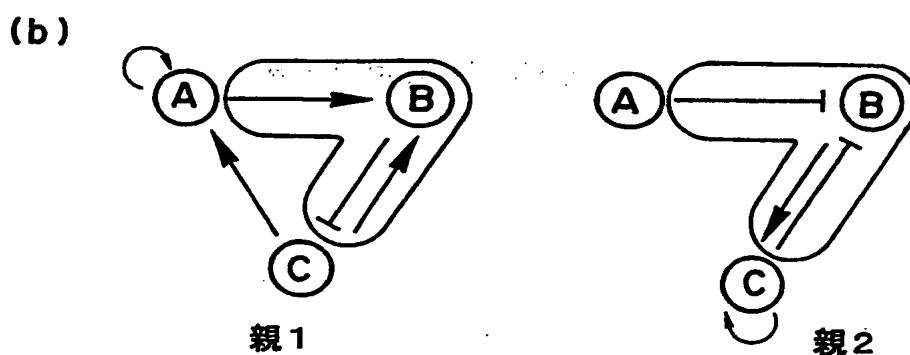
子 1

$$\begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}$$

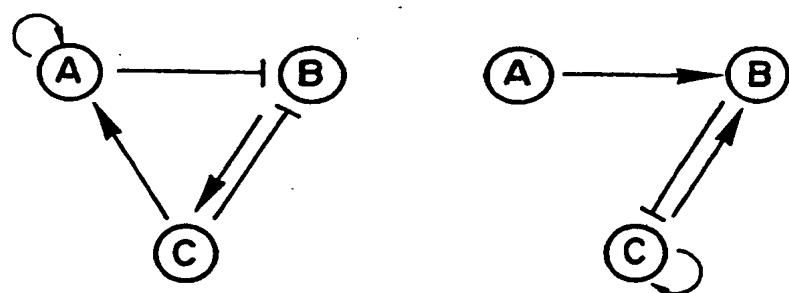
子 2

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 1 \end{pmatrix}$$

交叉



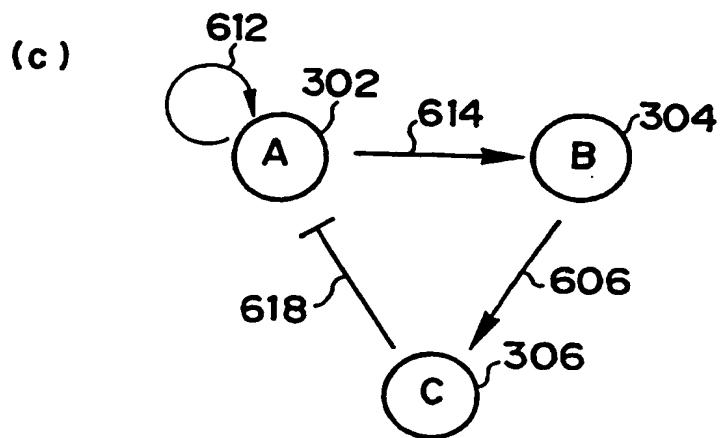
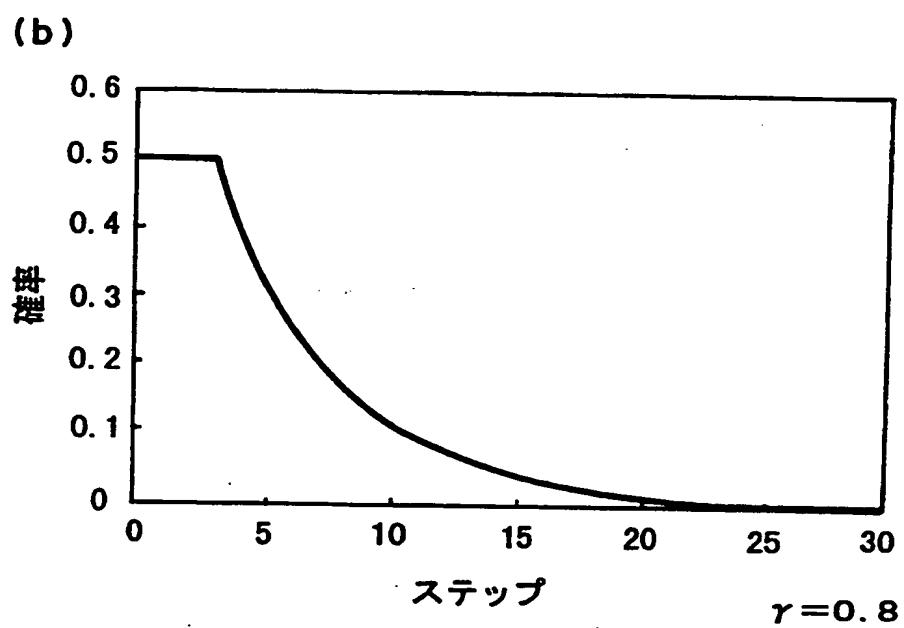
↓ 交叉



**THIS PAGE BLANK (USPTO)**

15 / 18

## 第 15 図

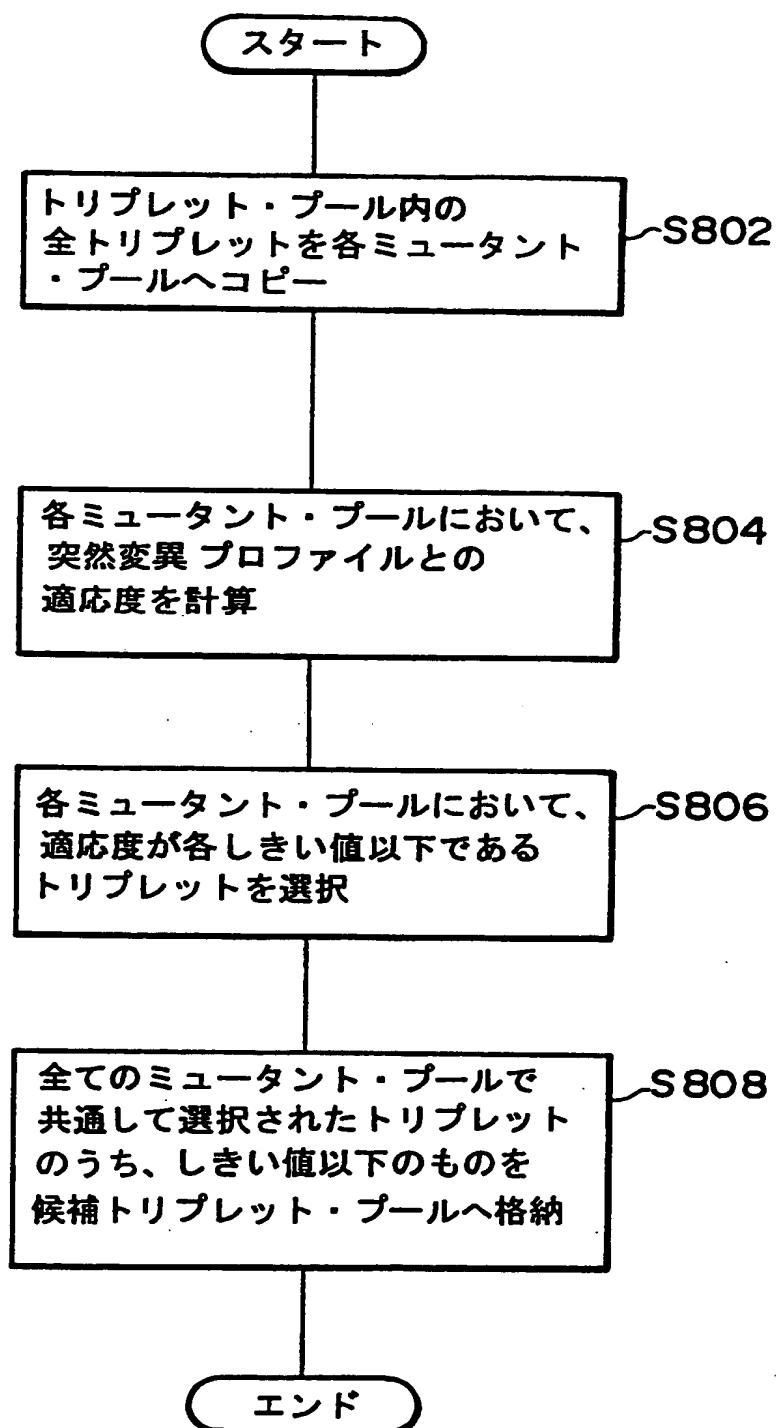


**HIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**

16 / 18

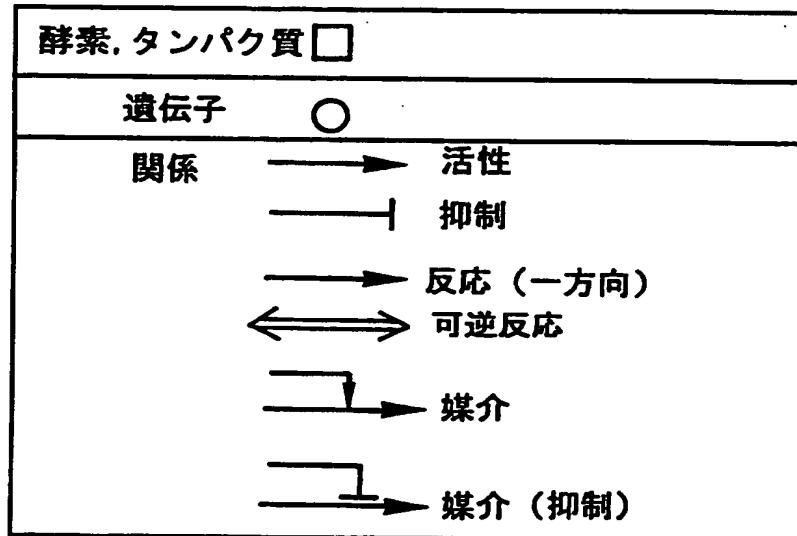
## 第 16 図



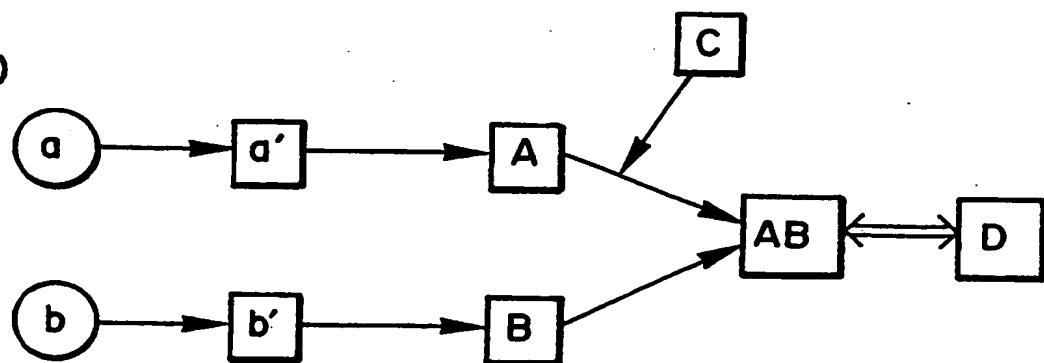
**THIS PAGE BLANK (USPTO)**

## 第 17 図

(a)



(b)

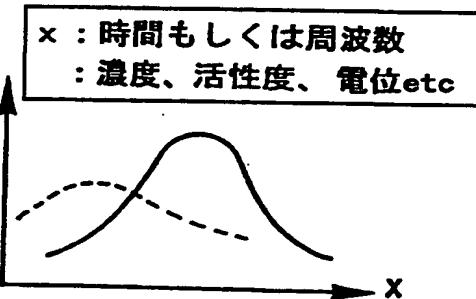


**THIS PAGE BLANK (USPTO)**

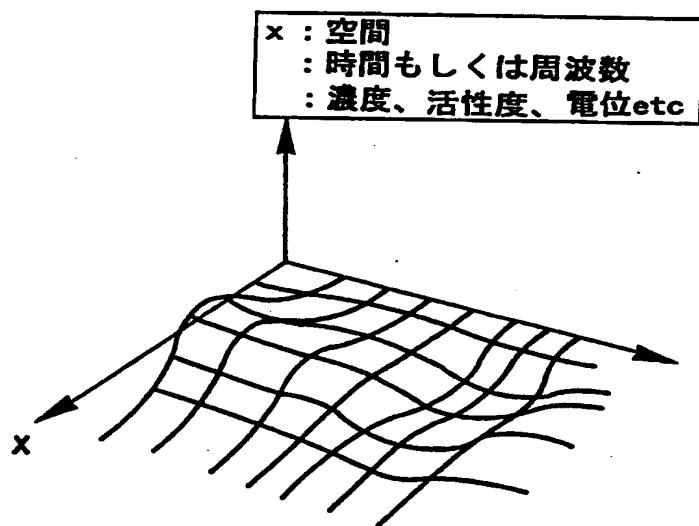
18 / 18

## 第 18 図

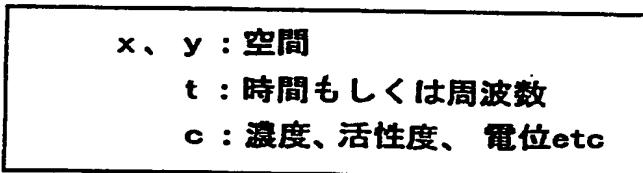
(a)



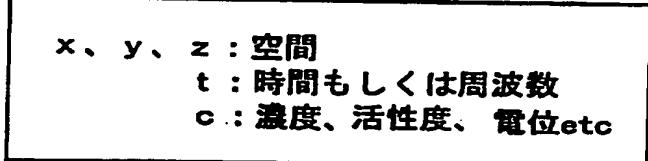
(b)



(c)



(d)



**THIS PAGE BLANK (USPTO)**

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP00/04050

**A. CLASSIFICATION OF SUBJECT MATTER**  
Int.Cl<sup>7</sup> G06N3/00

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

Int.Cl<sup>7</sup> G06N1/00-G06N3/12  
Int.Cl<sup>7</sup> G06F17/30  
Int.Cl<sup>7</sup> C12N15/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched  
Jitsuyo Shinan Koho 1926-1996 Jitsuyo Shinan Toroku Koho 1996-2000  
Kokai Jitsuyo Shinan Koho 1971-2000 Toroku Jitsuyo Shinan Koho 1994-2000

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
JICST DATA BASE (JOIS)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO, 90/11568, A1 (HONEYWELL INC.), 04 October, 1990 (04.10.90), Claims & US, 5140530, A & EP, 465489, B1 & JP, 2881711, B2	1-11
X	Tetsushi YATA et al., "Kakure Markov Model to Identeki Algorithm ni yoru DNA Hairetsu no Signal Pattern Chuushitsu", Transactions of Information Processing Society of Japan (IPSJ), Vol.37, No.6, 15 June, 1996 (15.06.96), pp.1117-1129	1-11
A	John Yen, et al., "A Hybrid Approach to Modeling Metabolic Systems Using a Genetic Algorithm and Simplex Method", IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS PART B: CYBERNETICS, Vol.28, No.2, April 1998 (30.04.98), pp.173-191	1-11

Further documents are listed in the continuation of Box C.

See patent family annex.

\* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search  
13 September, 2000 (13.09.00)

Date of mailing of the international search report  
26 September, 2000 (26.09.00)

Name and mailing address of the ISA/  
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

**THIS PAGE BLANK (USPTO)**

## A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int. C1' G06N3/00

## B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))

Int. C1' G06N1/00 - G06N3/12  
 Int. C1' G06F17/30  
 Int. C1' C12N15/00

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報 1926-1996年  
 日本国公開実用新案公報 1971-2000年  
 日本国実用新案登録公報 1996-2000年  
 日本国登録実用新案公報 1994-2000年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

JICSTデータベース (JOIS)

## C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
X	WO, 90/11568, A1 (HONEYWELL INC.) , 4. 10月. 1990 (04. 10. 90), 特許請求の範囲, & US, 51 40530, A & EP, 465489, B1 & JP, 28 81711, B2	1-11
X	矢田哲士他, 「隠れマルコフモデルと遺伝的アルゴリズムによるDNA配列のシグナルパターン抽出」, 情報処理学会論文誌, 第37卷, 第6号, 15. 6月. 1996年 (15. 06. 96), p. 1117-p. 1129	1-11

 C欄の続きにも文献が列挙されている。 パテントファミリーに関する別紙を参照。

## \* 引用文献のカテゴリー

「A」特に関連のある文献ではなく、一般的技術水準を示すもの  
 「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの  
 「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)  
 「O」口頭による開示、使用、展示等に言及する文献  
 「P」国際出願日前で、かつ優先権の主張の基礎となる出願

## の日の後に公表された文献

「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの  
 「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの  
 「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの  
 「&」同一パテントファミリー文献

国際調査を完了した日

13. 09. 00

国際調査報告の発送日

26.09.00

国際調査機関の名称及びあて先

日本国特許庁 (ISA/JP)  
 郵便番号 100-8915  
 東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

官司 卓佳



5M 9555

電話番号 03-3581-1101 内線 3597

C (続き) . 関連すると認められる文献		関連する 請求の範囲の番号
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	
A	John Yen, et al, 「A Hybrid Approach to Modeling Metabolic Systems Using a Genetic Algorithm and Simplex Method」, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS PART B:CYBERNETICS, Vol. 28, No. 2, April 1998 (30. 04. 98), p. 173-p. 191	1-11